UC San Diego

UC San Diego Previously Published Works

Title

Effect of natural genetic variation on enhancer selection and function.

Permalink

https://escholarship.org/uc/item/9nb2v54n

Journal

Nature, 503(7477)

ISSN

0028-0836

Authors

Heinz, S

Romanoski, CE

Benner, C

et al.

Publication Date

2013-11-01

DOI

10.1038/nature12615

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2014 May 28.

Published in final edited form as:

Nature. 2013 November 28; 503(7477): 487–492. doi:10.1038/nature12615.

Impact of natural genetic variation on enhancer selection and function

S. Heinz[≠], **C.E. Romanoski**[≠], **C. Benner**[≠,†,+], **K.A. Allison**[≠], **M.U. Kaikkonen**[≠,§], **L.D. Orozco**[¥], and **C.K. Glass**[≠,%,+,#]

[≠] Department of Cellular and Molecular Medicine, University of California, La Jolla, CA USA [%]Department of Medicine, University of California, La Jolla, CA USA [†] Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, La Jolla, CA USA [+]San Diego Center for Systems Biology, University of California, Los Angeles, USA [¥]Department of Human Genetics, University of California, Los Angeles, USA [§] Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland.

Abstract

The mechanisms by which genetic variation affects transcription regulation and phenotypes at the nucleotide level are incompletely understood. Here, we use natural genetic variation as an *in vivo* mutagenesis screen to assess the genome-wide effects of sequence variation on lineage-determining and signal-specific transcription factor binding, epigenomics, and transcriptional outcomes in primary macrophages from different mouse strains. We find substantial genetic evidence supporting the concept that lineage-determining transcription factors (LDTFs) define epigenetic and transcriptomic states by selecting enhancer-like regions in the genome in a collaborative fashion and facilitating binding of signal-dependent factors. This hierarchical model of transcription factor function suggests that limited sets of genomic data for LDTFs and informative histone modifications can be used for prioritization of disease-associated regulatory variants.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

[#] Correspondence: ckg@ucsd.edu.

SH, CER KAA and CKG: University of California, San Diego, Department of Cellular and Molecular Medicine, 9500 Gilman Drive, Mail Code 0651, La Jolla, CA 92093, USA

CB: Salk Institute, Integrative Genomics and Bioinformatics Core, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

MUK: University of Eastern Finland, Department of Biotechnology and Molecular Medicine, P.O. Box 1627, 70211 Kuopio, Finland

LDO: University of Californa, Los Angeles, Department of Molecular Cell and Developmental Biology, 3000 Terasaki Life Sciences Building, 610 Charles Young Drive East, Los Angeles, CA 90095, USA

Author Contributions

SH, CKG and CER designed the study; SH, CER, KAA, MUK and LDO performed experiments; CER performed all genetic variation-related analysis; CB wrote custom code for HOMER2 and analyzed data; KAA and SH analyzed data; CER, SH and CKG wrote the manuscript.

The authors declare no competing financial interests. Data is available in GEO accession GSE46494.

Supplementary Information

Supplementary Tables 1-3 and Data Source Files 1 and 2 accompany this manuscript.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Inter-individual genetic variation is a major cause of diversity in phenotypes and disease susceptibility. While sequence variants in gene promoters and protein-coding regions provide obvious prioritization of disease-causing variants, the majority (88%) of GWAS loci are in non-coding DNA, suggesting regulatory functions[1]. Prioritization of functional intergenic variants remains challenging due in part to an incomplete understanding of how regulation is achieved at the nucleotide level in different cell types and environmental contexts[2-11]. While recent studies have described important roles for lineage-determining transcription factors (LDTFs), also referred to as pioneer factors or master regulators, in selecting cell type-specific enhancers[12-15], the sequence determinants that guide their binding are poorly understood. Previous findings in macrophages and B cells suggest a hierarchical model of regulatory function[6], where a relatively small set of LDTFs collaboratively compete with nucleosomes to bind DNA in a cell type-specific manner (**Fig 1a**, i->ii). The binding of these factors is proposed to 'prime' DNA by initiating deposition of histone modifications that are associated with *cis*-active regulatory regions (**Fig. 1a**, ii -> iii) and enable concurrent or subsequent binding of signal-dependent transcription factors that direct regulated gene expression (**Fig. 1a**, iii ->iv)[6,13,15,16]. In principle, this model provides a straightforward framework that allows non-coding variants to be classified with respect to their ability to directly perturb LDTF binding and their potential to exert indirect effects on binding of other LDTFs and signal-dependent transcription factors. To test the validity of this model and its ability to explain effects of genetic variation on transcription factor binding and function, we exploited the naturally-occurring genetic variation between the inbred C57BL/6J and BALB/cJ mouse strains (~4 million SNPs and ~750k InDels[17]) as an '*in vivo* mutagenesis screen.'

Direct effects of genetic variation

First, we quantified genome-wide binding patterns of macrophage LDTFs PU.1 and C/EBPα from both mouse strains using ChIP-Seq. These experiments identified a combined 82,154 PU.1 and 54,874 C/EBPα peaks, with less than 1% of sites exhibiting highly significant strain-specific binding (PU.1, n=496; C/EBPα, n=263; 4-fold tag count ratio, FDR < $1e^{-14}$, >90% located >3 kb from gene promoters) (**Fig. 1b, c, Extended Data Fig. 1a**). Strain-specific binding was defined using biological ChIP-Seq replicates, which yielded <0.2% empirical false positives (**Extended Data Fig. 1b-g**). Differential binding of PU.1 and C/EBPα was significantly correlated with differential expression of the nearest gene as measured by RNA-Seq (**Fig. 1d**). There were no apparent differences in genomic context for strain-similar and strain-specific binding at inter- or intragenic sites (>3 kb to promoters) as defined by CpG content, distance from nearest gene or repetitive element, or conservation score (**Extended Data Fig. 2a**). Instead, strain-specific binding was highly correlated with polymorphism frequency. We observed 5-fold enrichment of polymorphisms at strain-specific versus strain-similar PU.1-bound and C/EBPα-bound regions (**Fig. 1e, Extended Data Fig. 2b)**, with the greatest variant density at the peak centers**, (Extended Data Fig. 2c,d)**.

To investigate direct effects of sequence variants on transcription factor binding, we identified the most enriched position weight matrices (PWM) in genomic regions marked by histone H3 lysine 4 di-methylation (H3K4me2) or bound by PU.1 or C/EBPα (**Extended**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data Fig. 3a, Supplementary Table 1). This analysis consistently identified consensus and degenerate motifs for the LDTFs PU.1, C/EBP and AP-1 as the most highly enriched PWMs. Notably, the frequency of mutations in these motifs increased with strain-specific binding of PU.1 and C/EBPα (**Extended Data Fig. 2e,f**). Excluding strain-specific loci without *cis*-variation (~11%), 41% of strain-specific PU.1 binding directly associated with strain-specific mutations in PU.1 motifs in the other strain. For C/EBPα, 44% of strain-specific binding associated with strain-specific C/EBPα motifs (**Fig. 1f**).

Although strain-specific binding of PU.1 and C/EBPα was highly linked to strain-specific motif mutations, strain-specific motif mutations were also associated with strain-similar binding (**Extended Data Fig. 3c,d**). This raised the question as to whether specific features of motif mutations could be used to predict strain-specific binding. Comparison of motif mutations in strain-specific to strain-similar peaks revealed three distinct attributes contributing to predictive power. First, mutated motifs within 20 bp of the experimentally defined binding centers were more highly associated with an effect on binding (PU.1, p = 1.6e-4; C/EBPα, p = 0.036; **Extended Data Fig. 4a-d**). Second, the presence of alternate motifs within 100 bp of the PU.1 peak centers significantly buffered the effect of strain-specific PU.1 motifs (**Extended Data Fig. 4e,f**). Third, after removing peaks with alternative motifs, analysis of the nucleotides mutated enabled delineation of an empirically defined functional motif that revealed a strong relationship between ‘core’ mutations and altered binding (**Fig. 1g, Extended Data Fig. 4g-i**, p=3.2e-8 PU.1 and p=5.1e-4 C/EBP). Taken together, core motif mutations <20 bp from the peak center that lacked alternative motifs were 3.5× and 3× more likely to occur in differential versus similar bound peaks for PU.1 and C/EBPα, respectively (**Extended Data Fig. 4j,k**). Notably, up to 90% of these mutations were located in differentially bound peaks (**Extended Data Fig. 4l,m**). To investigate the possibility that an algorithm incorporating these characteristics could be used to predict the impact of a specific motif mutation on transcription factor binding, we performed ChIP-Seq analysis for PU.1 in macrophages derived from a third inbred strain of mice, NOD/ShiLtJ (NOD). Of the ~1.4 million identifiable PU.1 motifs in the C57BL/6J reference genome, 18,322 contain SNPs that mutate the PU.1 motif in the NOD genome. 1.6% of these mutations were associated with strain specific binding (**Fig. 1h**). Of the 244 NOD PU.1 motif mutations located in PU.1-bound regions in C57BL/6J or BALBc, 68% were associated with strain-specific binding. When considering all three variables (motif distance, alternative motif, and motif core, **Extended Data Fig. 5**), 88% of the predicted functional mutations were consistent with impaired PU.1 binding in NOD (**Fig. 1h**).

Variation and collaborative LDTF binding

To investigate the potential impact of mutations in LDTF recognition motifs on collaborative binding, we analyzed all strain-specific PU.1 or C/EBPα binding events in regions containing LDTF motif mutations. PU.1 motif mutations resulting in loss of PU.1 binding were frequently associated with a corresponding loss of nearby C/EBPα binding in the absence of C/EBP motif mutations (**Fig. 2a**, top). Conversely, C/EBP motif mutations resulting in loss of C/EBPα binding were frequently associated with a corresponding loss of nearby PU.1 binding in the absence of PU.1 motif mutations (**Fig. 2a**, middle). Similar

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

results were observed at locations containing strain-specific mutations in AP-1 binding motifs, but intact PU.1 and C/EBP motifs (**Fig. 2a**, bottom).

We next considered the global relationships of mutations in PU.1, C/EBP, and AP-1 motifs with strain-specific binding of PU.1 and C/EBPα, taking into account both consensus and 'weak' motifs for PU.1 and C/EBP. NF-κB motifs were included as controls that were not expected to affect PU.1 or C/EBPα binding in unstimulated macrophages (**Extended Data Fig. 3a,b, Supplementary Table 1**). While mutations in PU.1 motifs had the strongest effect on strain-specific PU.1 binding, mutations exclusively in C/EBP and/or AP-1 motifs also significantly correlated with differential PU.1 binding relative to similarly bound loci (**Fig. 2b)**. Similar relationships were observed for C/EBP **(Extended Data Fig. 6a**). The motif distance distributions for co-bound factors were broad (half width ~100 nt), and only a minor subset of sites exhibited defined distances expected for direct protein-protein interactions **(Extended Data Fig. 6b**), suggesting transcription factor-nucleosome competition as the driving force behind the collaborative binding behavior[6,18]. Together, strain-specific mutations in nearby C/EBP and AP-1 motifs were associated with ~15% of strain-specific PU.1 binding at sites with strain-similar PU.1 motifs. Mutations in nearby PU.1 and AP-1 motifs were associated with ~30% of strain-specific C/EBPα binding at sites with strain-similar C/EBP motifs (**Fig. 1f**). Overall, 48% of strain-specific PU.1 binding and 57% of C/EBPα binding was associated with at least one assignable LDTF motif mutation (**Fig. 1f**). To genetically test whether these correlations are consistent with a collaborative binding model, we considered all LDTF motif mutations and evaluated their effects on PU.1 binding in macrophages derived from NOD mice. For polymorphic strain-specific PU.1 loci containing strain-specific LDTF motifs (n = 220), PU.1 binding profiles matched the strain with shared alleles for 91% and 92% of cases (**Fig. 3a**). At 8% (n = 17) of the loci, the NOD genome broke the C57BL/6J/BALB/cJ haplotypes, and in all cases, the NOD genotype at the LDTF motif variant matched the strain with similar binding (**Supplementary Table 2**), indicating that these variants are likely the cause of binding differences. An example is shown in **Fig. 3b**, where PU.1 binds in C57BL/6J but not in BALB/cJ or NOD. Only one SNP in this region is associated with PU.1 binding exclusively in C57BL/6J; here, the T allele forms part of a neighboring AP-1 motif in C57BL/6J that is mutated by the C allele in BALB/cJ and NOD. These findings provide genetic evidence that PU.1 binding to this location is dependent on collaborative interactions with AP-1.

To confirm that the allele-specific binding also occurs in heterozygous cells, we performed ChIP-Seq for PU.1 and C/EBPα in macrophages from CB6F1/J hybrid mice, which are F1 offspring of a C57/BL/6J × BALB/cJ cross. In the great majority of cases, alleles bound specifically in a parental strain were also bound preferentially in the F1 generation. (**Fig. 3c** and **Extended Data Fig. 6c**).

Given the genetic evidence that LDTFs collaborate to bind DNA, we next tested the extent to which strain-specific LDTF binding explained promoter-distal (>3 kb) strain-specific histone modification events, such as H3K4me2 and H3K27Ac deposition, which respectively mark 'primed' and 'active' chromatin[19,20] (**Fig. 1a**, ii->iii). Genomic regions exhibiting strain-specific binding of PU.1 and C/EBPα were associated with strain-specific H3K4me2 and H3K27ac (**Fig. 2a**, right columns). Strain-specific histone modifications

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

correlated with nearby gene expression (**Fig. 2c**), and H3K27Ac modification tracked with the corresponding parental allele in CB6F1/J hybrid mice (**Fig. 3d**). Strain-specific binding of PU.1 and C/EBPα were individually correlated with H3Kme2 and H3K27Ac deposition, with the combined binding of both factors exhibiting even greater correlation than the individual factors (**Extended Data Fig. 7a-f**). Further, LDTF motif mutations segregated with differential LDTF binding and histone modifications (**Fig. 2d, Extended Data Fig. 7g**). Together, these findings support the concept that LDTFs play quantitatively important roles in establishing these histone modifications, likely through initiating transcription in a combinatorial fashion[21].

Expression quantitative trait loci (eQTLs) are polymorphic loci whose alleles are associated with individual RNA expression levels across a population[22]. Thus, eQTLs define active gene regulatory loci and provide an alternative method for assigning regulatory function to gene expression. To interrogate the relationship between histone modification and eQTLs, we analyzed previously reported eQTL data from 85 inbred mouse strains in the Hybrid Mouse Diversity Panel (HMDP) in primary macrophages[23] (see Methods). We found that eQTLs overlapped H3K4me2- or H3K27Ac-marked regions at frequencies greater than expected by chance, supporting the role of histone modifications as landmarks of regulatory activity (hypergeometric test p-values: H3K4me2 = $1e^{-2147}$, H3K27Ac = $1e^{-2290}$). Next, given the highly cell type-specific nature of gene regulation[24], we hypothesized that eQTLs from different cell types would be reflected in the histone modification profiles in the same cell type. We examined liver and macrophage eQTLs for a set of ~130k SNPs from the HMDP[25] for overlap with H3K27Ac loci defined in macrophages or in liver, pro-B, or mouse ES cells[20]. Macrophage eQTL were more significantly enriched for overlap with macrophage H3K27Ac regions than liver H3K27Ac. Similarly, liver eQTL were most significantly enriched with liver H3K27Ac relative to macrophage H3K27Ac (**Fig. 2e**). Clustering of H3K27Ac profiles revealed that liver and ES H3K27Ac profiles are most similar (**Extended Data Fig. 7h**), providing an explanation as to why liver eQTLs were highly enriched in mES H3K27Ac regions.

LDTF motif mutations affect NF-κB binding

To evaluate the prediction that primed regulatory loci (containing H3K4me2) often require additional binding of signal-dependent TFs to achieve regulatory activity (**Fig 1a**, iii->v), we treated C57BL/6J and BALB/cJ macrophages with Kdo2-Lipid A (KLA), a potent and specific agonist of TLR4[26]. KLA treatment causes NF-κB to enter the nucleus, bind DNA and regulate several hundred target genes[26,27]. We performed ChIP-Seq for PU.1, C/EBPα and the RelA/p65 component of NF-κB in untreated and KLA-treated macrophages and observed that 61% of sites that gained p65 were pre-bound by PU.1 and/or C/EBPα without KLA. *De novo* motif analysis indicated that an AP-1 motif was present in 42% of the remaining sites, suggesting that AP-1 is responsible for priming a large proportion of the p65 cistrome (**Extended Data Fig. 8a**), in line with previous reports[16].

To further interrogate the dependence of p65 on LDTFs we focused on sites that gained p65 only in one strain (n = 932, >90% promoter-distal, **Extended Data Fig. 1a**, **Fig. 4a**, 4th column). In the vast majority of cases, PU.1 and/or C/EBPα were bound prior to KLA

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

treatment only in the strain exhibiting p65 binding (**Fig. 4a**). In addition, strain-specific p65 binding primarily occurred at loci already marked by H3K4me2, and led to an increase of H3K27Ac, consistent with the proposed model. To analyze the effects of genetic variation on transcription factor motifs, we performed strain-specific LDTF and NF-κB motif finding in polymorphic strain-specific p65-bound peaks (n = 750) (**Extended Data Fig. 3b**). Notably, p65 binding was influenced by mutations in individual LDTF motifs to a similar extent as mutations in the NF-κB motif itself (**Fig. 4b, Extended Data Fig. 8b**). For strain-specific p65 binding events, 34% could be attributed to assignable mutations in PU.1, C/EBP, or AP-1 motifs, whereas 9% could be explained by mutations in the assignable NF-κB motifs themselves (**Fig. 4c**). RelA/p65 is known to bind to degenerate and non-canonical motifs[28] that might not be captured by *de novo* motif analysis. To gain motif-independent insight into variant location and strain-specific TF binding, we assessed the variant frequency relative to the centers of strain-specific p65 peaks. Similar to strain-specific PU.1 and C/EBPα peaks, strain-specific p65 peaks are in regions of higher variant density than strain-similar peaks (**Extended Data Fig. 8c**). In contrast to LDTFs, where strain-specifically bound regions have a high variant density at their peak centers, the distribution of variants at strain-specific p65 peaks is significantly different from those of the LDTFs (Kolmogorov-Smirnov p-value < 0.013) as it contains fewer variants at the peak centers and is broader (**Fig. 4d, Extended Data Fig. 8d-f**). This is consistent with p65 binding being more affected by sequence variation in motifs of neighboring factors than LDTFs.

Overall, strain-specific p65-bound regulatory sites were significantly correlated with nearby genic transcription and mRNA production (**Fig. 4e**). We tested strain-specifically bound and epigenetically marked putative enhancer sequences with strain-specific mutations for differential enhancer function in transient and stable reporter assays (**Fig. 5a, Extended Data Figs. 9a, b**). We observed the predicted strain-specific enhancer activity for 18 of 20 of these genomic sequences. Conversely, enhancer elements with sequence variation in non-core nucleotides that were not predicted to alter PU.1 or C/EBP binding and that exhibited strain-similar binding patterns exhibited similar enhancer activity (**Extended Data Fig. 10a**).

Lastly, we tested whether the predicted motif-disrupting variants could specifically explain strain-specific enhancer activity by swapping variants at the putative causative alleles in C57BL/6J to BALB/cJ while maintaining the genetic background for the remainder of the enhancer sequences. Representative examples in which reversal of such SNPs in PU.1, C/EBP and p65 motifs reversed strain-specific enhancer activity are illustrated in **Fig. 5c and Extended Data Figs. 10b,c**. In contrast, reversal of nearby SNPs not predicted to alter LDTF motifs had no effect on strain-specific enhancer activity (**Extended Data Fig. 10c**).

Discussion

In concert, we have exploited natural genetic variation to test a collaborative model for enhancer selection and function, and conversely explored the ability of this model to explain strain-specific differences in transcription factor binding and epigenetic features associated with functional enhancers in macrophages. These studies provide genetic evidence that lineage-determining transcription factors are dependent on collaborative binding to variably

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

spaced DNA recognition motifs in order to select enhancers and enable binding of signal-dependent transcription factors. Notably, the variable motif distances observed at co-bound LDTF loci suggests that collaborative binding does not generally require direct protein-protein interactions between the involved transcription factors. The proposed hierarchical LDTF collaborative model provides a conceptual framework for prioritization of non-coding disease-associated regulatory variants. While all cells express hundreds of transcription factors, a large fraction of functional enhancers (~70% in macrophages) are characterized by collaborative interactions involving relatively small sets of lineage determining transcription factors (e.g., PU.1, AP-1 and C/EBPs). The requirement for collaborative binding interactions provides an explanation for why transcription factor binding is lost at sites where mutations do not occur in the cognate recognition motif. In the case of NF-κB, for example, mutations in the motifs for lineage determining factors were approximately three times more likely to result in decreased binding of NF-κB than mutations in the NF-κB binding site itself. An essential step in leveraging the collaborative model to pinpoint potential disease-causing variants is the definition of relevant LDTF binding sites and functionally important variants. At the current level of genome annotation, this cannot be achieved by analysis of DNA sequence alone. For example, there are ~1-2×10^6 identifiable PU.1 binding sites in the human[29] and mouse genomes, but <10% are actually occupied by PU.1 in macrophages. By experimentally defining strain-similar and strain-specific binding patterns for PU.1, the relevant sites at which mutations can result in altered function are identified. Comparison of PU.1 motif mutations associated with strain-specific versus strain-similar binding allowed the genetic definition of a functional binding matrix and additional distinguishing features that enabled accurate prediction of functional mutations in a third strain. Thus, by collecting a relatively limited set of genomic binding data for LDTFs and informative histone modifications, this analytical approach can be exploited to explain a greater extent of variation in enhancer selection and function than previously possible[7,10]. To further increase the specificity and sensitivity for detecting functional variations, identification of transcription factor motifs that permit binding but diverge from the consensus PWM, i.e. “weak” motifs, needs to be improved, as such sites are more likely to be affected by mutation[29,30]. In addition, transcription factors less abundant than LDTFs likely play individually small but collectively significant roles. At a larger scale, non-*cis*-acting, long-range epigenetic mechanisms may also be important for enhancer selection. A major goal for the future will be to extend these approaches to understanding natural genetic variation associated with human disease.

Methods

Animals and Cell Culture

Thioglycolate-elicited peritoneal macrophages were collected 4 days post-injection from male 6-8 wk C57BL/6J, BALB/cJ or CB6F1/J Hybrid mice and plated at 20 × 10^6 cells/15-cm petri dish in RPMI1640 + 10% FBS + 1X PenStrep. One day after plating, cells were treated with fresh media with or without 100 ng/ml Kdo2-LipidA (KLA) for 1 hour and then directly used for downstream analyses. All animal experiments were performed in compliance with the ethical standards set forth by UC San Diego's Institutional Annual Care and Use Committee (IUCAC).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ChIP-Seq and Feature Identification

Media was decanted and cells were fixed at room temperature with either 1% formaldehyde/PBS for 10 minutes (for PU.1, C/EBPa, H3K27Ac ChIPs) or 2 mM disuccinimidylglutarate (DSG, Pierce)/10% DMSO/PBS for 30 minutes followed by 1% formaldehyde/PBS for another 15 minutes (p65). After quenching the reaction by adding glycine to 0.125 M, cells were washed twice with PBS and snap-frozen in dry-ice/methanol. ChIPs for PU.1 (Santa Cruz, sc-352), C/EBPa (Santa Cruz, sc-61) were performed exactly as described previously[6]. The H3K27Ac (Abcam, ab4729) ChIP was performed in the presence of 1 mM butyric acid. For p65 (Santa Cruz, sc-372), IP conditions were identical to the ones described before[6], except that pre-clearing was omitted, and the ChIP was performed with 5 μg antibody (Santa Cruz, sc-372) pre-bound to 50 μl Protein A Dynabeads (Invitrogen) for 30 minutes in 0.5% BSA/TE, and a final wash with TE/50 mM NaCl was performed before elution. ChIP-Seq library preps for the initial p65 ChIPs were performed as described before[6], libraries for the replicates were prepared using magnetic beads similar to previously described procedures[12]. ChIPs for H3K4me2 were carried out on MNase-digested chromatin as described previously[31]. To control for open chromatin and library biases, input chromatin libraries after sonication were sequenced for each strain, crosslinking condition and ChIP lysis protocol. Sequencing libraries were prepared as previously described[6] using barcoded adapters (NextFlex, Bioo Scientific), and sequenced for 50 cycles on a Hi-Seq 2000 (Illumina) using CASAVA1.7 or 1.8.

C57BL/6J and BALB/cJ demultiplexed fastq files were mapped to both the mm9 reference (C57BL/6J) genome and the BALB/cJ contigs[17] using Bowtie0.12.7[33] with the options –m 1 --best –n 3 -e 200. Mapping parameters for C57BL/6J and BALB/cJ data allowed 3 mismatches in the 28 bp seed sequence with up to 5 high quality mismatches in the entire 50 bp read. NOD ChIP-Seq data were mapped to the mm9 genome using the above options. To identify allele-specific reads from CB6F1/J data, ChIP-seq reads were aligned to the C57BL/6J or BALB/cJ sequence using Bowtie2-2.0.0-beta7[34] allowing 0 mismatches in 32 bp reads with options -N 0 -L 32 --score-min L,0,0 --gbar 17. Tags mapping to both genomes were discarded. Resulting allele-specific reads were counted for regions of interest.

For C57BL/6J and BALB/cJ data, reads mapping to only one genome were discarded (<2% of total) to avoid bias caused by mappability differences, and reads mapping to both were assigned to the mm9 genomic location. Genomic binding peaks for transcription factors PU.1, C/EBPa, and p65, were identified using the *findPeaks* command in the HOMER (http://biowhat.ucsd.edu/homer/) software suite[6] with default settings of *–style factor*: 200 bp peaks with 4-fold tag enrichment and 0.001 FDR significance over background (ChIP input), 4-fold enrichment over local tags, and normalization to 10 million mapped tags per experiment. H3K4me2 and H3K27Ac regions used for initial *de novo* motif finding (Extended Data Fig. 3a) were identified using the default parameters of *findPeaks –style histone* with the addition of –nfr centering for H3K4me2 MNase data. For H3K4me2 and H3K27Ac peaks identified for comparison to LDTF binding and mutation events (e.g., **Fig. 2d**), findPeaks –style peaks was used to define more focal, non-gene associated loci. In particular, H3K27Ac regions were identified with the *findPeaks* options *–style factor* using –size 1000bp –L2 (2-fold enrichment over local tags). H3K27Ac peaks were merged between

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

strains using *mergePeaks –size given* and peaks were resized to 1 kb. Peaks within 3 kb of gene promoters were excluded from further analysis. H3K4me2 peaks were identified using *findPeaks* options –style factor –size 500 –L2 –C0 (which allows for unlimited tags considered per genome position as may occur with MNase data). Peaks were then centered on the best nucleosome-free region (nfr) within a 1 kb window using *getPeakTags –nfr*. Peak files between strains were merged with *mergePeaks –size given* and H3K4me2 tags were counted in 1 kb regions centered on the merged peak file definitions. Peaks were then re-centered based on the best nfr in 1 kb identified by *getPeakTags –nfr* according to the strain with more H3K4me2 tags. Peaks were extended to 1 kb and restricted to those >3 kb from gene promoters.

Strain-Specific Feature and Motif Identification

To determine strain-specific binding and epigenetic modification events we counted the number of sequencing tags (normalized to 10 million) at peaks/regions identified in the set of combined peaks/regions from both strains. We normalized the mean peak tag count to be equal in each strain and compared the tag counts in each region and required strain-specific peaks/regions to exhibit ≥4-fold difference in tag counts and an adjusted cumulative Poisson p-value corresponding to FDR<1e-14 [35] These criteria were based on empirical data relating replicate ChIP-seq experiments (**Extended Data Fig. 1b**). Individual genome sequences for C57BL/6J and BALB/cJ were constructed in regions of interest using the reference (C57BL/6J) sequence and replacing BALB/cJ alleles at SNPs and InDels reported in the vcf files from Keane *et. al.*[17].

Strain-Specific Motifs

De novo motif finding in ChIP-Seq-enriched regions from both mouse strains was used to define position weight matrices (PWM) for transcription factors of interest (**Extended Data Fig. 3a,b, Supplementary Table 1**). These PWM were used to define strain-specific motifs by using the options *homer2 –find <individual genome sequence>* in HOMER for each genome sequence for the regions of interest. The positions of the identified motifs were compared between strains taking into account shifts cause by InDels relative to peak start coordinates and which DNA strand matched the identified motifs. Motifs with alignments only in one genome were considered strain specific.

PolyA RNA-Seq

For each condition, RNA was isolated from 5×10^6 thioglycolate-elicited macrophages with Trizol LS, and 15 μg RNA were DNase-treated using TURBO DNase (Ambion) according to the manufacturer's instructions and ethanol-precipitated. PolyA-RNA was selected from 7 μg total RNA using the MicroPoly(A)Purist kit (Ambion), according to the manufacturer's instructions. The isolated RNA was hydrolyzed in 20 μl total volume with 2 μl RNA fragmentation buffer (Ambion) for 10 minutes at 70°C, the reaction stopped with stop buffer, and buffer was exchanged to Tris, pH 8.5 using P30 size exclusion columns (Bio-Rad). The fragmented RNA (30 ng) was 5'-decapped in 21 μl total volume containing 0.5 μl tobacco acid pyrophosphatase (TAP, Epicentre), 2 μl 10× TAP buffer, 1 μl SUPERase-IN, incubate for 2 h at 37°C. To dephosphorylate RNA 3' ends, 0.5 μl 10× TAP buffer, 1.5 μl

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

water, 0.5 μl 0.25 M MgCl2 (4.17 mM final –1 mM EDTA for maximum phosphatase activity), 0.5 μl 10 mM ATP (0.2 μM final to protect PNK) where added, and the reaction incubated with 1 μl PNK (Enzymatics) for 50 minutes at 37°C. RNA fragments were 5'-phosphorylated by adding 10 μl 10× T4 DNA ligase buffer, 63 μl water, 2 μl PNK, and 60 minute incubation at 37°C. RNA fragments were isolated using Trizol LS, precipitated in the presence of 300 mM NaAc and 2 μl Glycoblue (Ambion), washed twice with 80% ethanol and dissolved in 4.5 μl water.

To prepare sequencing libraries, 0.5 μl 9 μM of a 5'-adenylated sRNA3'MPX adapter / 5Phos/AG ATC GGA AGA GCA CAC GTC TGA /3AmMO/ (IDT, desalted; adenylated with Mth ligase (NEB) according to the manufacturer's instructions, phenol-chloroform/ chloroform-extracted, ethanol-precipitated with glycogen and dissolved in water at 9 μM) were heat-denatured together with the RNA for 2 minutes at 70°C, and ligated with 100 U truncated T4RNA ligase 2 K227Q (NEB) in 10 μl 1× T4 RNA ligase buffer without ATP, containing 10 U SUPERase-In and 15% PEG8000 for 2 hours at 16°C. To reduce adapter dimer formation, 0.5 μl 10 μM MPX_RT primer 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3' (IDT, desalted) was added and annealed to the ligation product by incubating at 75°C for 2 minutes, then 37°C for 30 minutes, then 25°C for 15 minutes. Finally, 0.5 μl 5 μM hybrid DNA/RNA sRNA5'h adapter 5'-GTT CAG AGT TCT ACA rGrUrC rCrGrA rCrGrA rUrC-3' (IDT) were ligated to previously capped RNA 5' ends by adding 2 μl T4 RNA ligase buffer, 6 μl 50% PEG8000 (15% final), 1 μl 10 mM ATP, 9.5 μl water and 0.5 μl (5 U) T4 RNA ligase 1 for 90 minutes at 20°C. To 15 μl ligation reaction, an additional 0.5 μl 10 μM MPX_RT primer were added, reactions were denatured at to 70°C for 1 minute, then placed on ice. RNA was reverse-transcribed by adding 3 μl 10× first strand buffer, 4.5 μl water, 1.5 μl 10 mM dNTP, 3 μl 0.1 M DTT, 1.5 μl RNaseOUT and 1 μl Superscript III reverse transcriptase (Invitrogen), and incubating for 30 minutes at at 50°C. Complementary DNA was isolated by adding 35 μl AMPure XL beads (Beckman), binding and washing according to manufacturer's instructions and dissolved in 40 μl TET (0.1 % Tween 20/TE). Libraries were PCR-amplified for 9 (polyA RNA-Seq), 11 (5'-GRO-Seq), 12 (rRNA-5'-RNA-Seq) or 13 (polyA-5'-RNA-Seq) cycles with 0.75 μM primers oNTI201 and TruSeq-compatible indexed primers (e.g. 5'-CAA GCA GAA GAC GGC ATA CGA GAT nnn nnn GTG ACT GGA GTT CAG ACG TGT GCT CTT-3' (desalted, IDT, index in lowercase letters) using Phusion Hot Start II (Thermo Scientific) in HF buffer containing 0.5 M betaine (98°C, 30s/12x(98°C, 10s/57°C, 25s/72°C, 20s)/ 72°C, 1min/4°C, ∞), and 175-225 bp fragments were size-selected on 10% PAGE gels. Libraries were diluted 1:10^5 with TET and quantified relative to samples of known cluster density by SYBR green Q-PCR with primers Solexa_1G_A 5'-AAT GAT ACG GCG ACC ACC GA-3', Solexa_1G_B 5'-CAA GCA GAA GAC GGC ATA CGA-3' (95°C, 15 min/25x(95°C, 10s/ 60°C, 60 s)) and sequenced for 51 (insert) +7 (index) cycles on a HiSeq 2000 sequencer (Illumina) with small RNA sequencing primer 5'-CGA CAG GTT CAG AGT TCT ACA GTC CGA CGA TC-3' and TruSeq Index sequencing primer (Illumina).

GRO-Seq

GRO-Seq was performed as described previously[32] using 10^7 cells per condition. RNA at RefSeq transcripts was quantified for GRO-Seq and PolyA-RNA-Seq by counting the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

normalized tags (to 10 million tags/experiment) in annotated exons for each RefSeq transcript.

Odds Ratio Calculations and Statistical Testing

Odds ratios for observing C57BL/6J-specific motif mutations relative to BALB/cJ-specific motif mutations in different classes of bound/modified loci (e.g., **Fig. 2b**) were calculated using $(p_1 / (1-p_1)) / (p_2 / (1-p_2))$ where p_1 is the frequency of C57BL/6J-specific motifs and p_2 is the frequency of BALB/cJ-specific motifs. For **Extended Data Fig. 4j,k**, p_1 is the frequency of indicated events occurring in differentially bound loci and p_2 is the frequency in similarly bound loci. Unless otherwise indicated, t-tests were two-sided assuming unequal variance.

eQTL Analysis

eQTL analysis was performed as previously described[23,25]. In brief, thioglycolate-elicited peritoneal macrophages were collected from 85 strains of mice. RNA was processed and hybridized to Affymetrix Genome HT_MG-430A. There were 22,416 probe sets analyzed after removing individual probes overlapping SNPs and probe sets with 8 or more probes overlapping SNPs. Expression data was RMA normalized.

3,918,755 SNPs with a minimum minor allele frequency of 10% originating from mouse Perlegen variation dataset[36] were imputed across the strains[37] and filtered to 3,695,041 SNPs based on proximity (<2 Mb) to transcription start sites of transcripts detectable by the microarray. Gene expression for each transcript was associated to SNPs within 2 Mb using the efficient mixed-model association (EMMA) mapping that corrects for population structure[38]. Association p-values less than 1×10^{-5} (<1% FDR) were deemed significant[23]. The 3,695,041 SNPs used for association mapping were overlapped with H3K4me2 and H3K27Ac regions. Since H3K4me2 and H3K27Ac regions ranged from 500-1500 bp whereas haplotype blocks averaged 300 kb, we considered SNPs outside H3K4me2/H3K27Ac regions that were in linkage disequilibrium (LD) with a SNP in H3K4me2/H3K27Ac regions as overlapping. Haplotype Blocks were estimated in Haploview[39] using 143 strains with the following options: blockMAFThresh=0.1, blockCutLowCI=0.8, blockCutHighCI=0.98, blockRecHighCI=0.9, blockInformFrac=0.95. SNPs in LD with Enhancer SNP were considered markers of H3K4me2 regions. To test for enrichment of significant eQTL in H3K4me2 regions, we used the Hypergeometric Distribution Function as follows:

$$P\,(X{=}k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

k successes = # significant eQTL in (or in LD with) H3K4me2 regions; m = # SNPs with significant eQTL, N = # total SNPs, n = # total SNPs in H3K4me2 regions.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Macrophage eQTL enrichment in enhancers from other cell types

The short read archive files were downloaded from GEO Accession GSE24164[20] for ChIP-sequencing for the H3K27Ac mark in whole liver (SRX027340), pro-B cells (SRX027345), and ES cells (SRX027331, SRX027332) and input chromatin as background (liver: SRX027343, pro-B: SRX027348, ES: SRX027352). Sequencing reads were mapped to the C57BL/6J genome. H3K27Ac regions were identified where tag pileups exceeded 4X the input tags using HOMER[6] and interrogated for enrichment of significant macrophage eQTLs as described for macrophage H3K4me2 and H3K27Ac regions above.

Reporter Assays and Mutation Analysis

One kb enhancers were PCR-amplified from C57BL/6J and BALB/cJ genomic DNA using genomic primers not overlapping variants that introduced terminal BamHI, BglII or BclI sites on one end and SalI or XhoI sites on the other end of the PCR products, depending on the restriction site content of the enhancer. These were digested with the respective restriction enzymes and ligated into a modified, BamHI and SalI-digested pGL4.10 luciferase reporter plasmid (Invitrogen) containing a minimal HSV-TK promoter derived from pTAL-Luc (Clontech) (see **Fig. 5a**). Alternatively, 1 kb fragments were amplified using primers that introduced overhangs identical to the sequences flanking the BamHI/SalI tandem site in the pGL4.10 plasmid. Fragments were purified from the PCR reaction by SPRI using magnetic beads and cloned into the BamHI/SalI-cut reporter plasmid described above using Gibson Assembly master mix (NEB) according to the manufacturer's instructions. Mutations were introduced by PCR amplification with complementary primers containing the mutation to be introduced in the center, followed by DpnI digest of the template and transformation of bacteria. All constructs were confirmed by sequencing. For each reporter assay, 300 ng plasmid was transfected into RAW264.7 macrophages using SuperFect (Qiagen) together with 300 ng UB6 promoter-driven beta-galactosidase reporter for transfection normalization in 24-well plates seeded with 1×10^5 cells 24 hours prior to transfection. 24 hours post-transfection, media alone (RPMI + 10%FBS) or media also containing 100 ng/ml KLA was added for an additional 16 hours. Luciferase activity was measured 24 hr post transfection using a Veritas microplate luminometer (Turner Biosystems) and normalized to beta-galactosidase activity (Applied Biosystem) for transfection efficiency. Each experiment was performed at least three independent times, with each reaction done in triplicates. Data represented mean ± s.d., and statistical significance was determined by one-sided t-test.

Stable transfected cell lines were made by transient co-transfection of the linearized reporter plasmids together with linearized neomycin resistance-expressing pcDNA3 vector as described above, followed by incubation with 275 μg/ml Geneticin (G418 Sulfate, Invitrogen) for 2-3 weeks. Bulk cells from stably transfected colonies were tested for luciferase activity and normalized to total protein concentration (DC Protein Assay, BioRad).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgements

We thank Aldons J. Lusis, Ph.D. (UCLA) for providing access to eQTL data (http://systems.genetics.ucla.edu/) and for productive conversations. We thank our Referees for highly constructive comments. We thank Daniel Pollard (UCSD) for discussions and suggestions, and Lynn Bautista for assistance with figure preparation. These studies were supported by NIH grants DK091183, CA17390 and DK063491 (C.K.G.). M.U.K. was supported by the Foundation Leducq Career Development award and grants from Academy of Finland, Finnish Foundation for Cardiovascular Research and Finnish Cultural Foundation, North Savo Regional fund. C.E.R. was supported by the American Heart Association Western States Affiliates (12POST11760017) and NIH (5T32DK007494).

References

1. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:9362–9367. doi:10.1073/pnas.0903103106. [PubMed: 19474294]
2. Cowper-Sal lari R, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat Genet. 2012; 44:1191–1198. doi:10.1038/ng.2416. [PubMed: 23001124]
3. Degner JF, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012; 482:390–394. doi:10.1038/nature10808. [PubMed: 22307276]
4. Gaffney DJ, et al. Dissecting the regulatory architecture of gene expression QTLs. Genome biology. 2012; 13:R7. doi:10.1186/gb-2012-13-1-r7. [PubMed: 22293038]
5. Gaulton KJ, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010; 42:255–259. doi:10.1038/ng.530. [PubMed: 20118932]
6. Heinz S, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell. 2010; 38:576–589. [PubMed: 20513432]
7. Kasowski M, et al. Variation in transcription factor binding among humans. Science. 2010; 328:232–235. doi:10.1126/science.1183621. [PubMed: 20299548]
8. Maurano MT, Wang H, Kutyavin T, Stamatoyannopoulos JA. Widespread site-dependent buffering of human regulatory polymorphism. PLoS genetics. 2012; 8:e1002599. doi:10.1371/journal.pgen.1002599. [PubMed: 22457641]
9. McDaniell R, et al. Heritable individual-specific and allele-specific chromatin signatures in humans. Science. 2010; 328:235–239. doi:10.1126/science.1184655. [PubMed: 20299549]
10. Reddy TE, et al. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. Genome research. 2012; 22:860–869. doi:10.1101/gr.131201.111. [PubMed: 22300769]
11. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome research. 2012; 22:1748–1759. doi:10.1101/gr.136127.111. [PubMed: 22955986]
12. Garber M, et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol Cell. 2012; 47:810–822. doi:10.1016/j.molcel.2012.07.030. [PubMed: 22940246]
13. Mullen AC, et al. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell. 2011; 147:565–576. doi:10.1016/j.cell.2011.08.050. [PubMed: 22036565]
14. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell. 2012; 151:994–1004. doi:10.1016/j.cell.2012.09.045. [PubMed: 23159369]
15. Trompouki E, et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell. 2011; 147:577–589. doi:10.1016/j.cell.2011.09.044. [PubMed: 22036566]
16. Ghisletti S, et al. Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. Immunity. 2010; 32:317–328. [PubMed: 20206554]
17. Keane TM, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011; 477:289–294. doi:10.1038/nature10413. [PubMed: 21921910]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

18. Mirny LA. Nucleosome-mediated cooperativity between transcription factors. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:22534–22539. doi:10.1073/pnas.0913805107. [PubMed: 21149679]
19. He HH, et al. Nucleosome dynamics define transcriptional enhancers. Nat Genet. 2010; 42:343–347. doi:10.1038/ng.545. [PubMed: 20208536]
20. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:21931–21936. doi:10.1073/pnas.1016071107. [PubMed: 21106759]
21. Kaikkonen MU, et al. Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. Mol Cell. 2013; 51:310–325. doi:10.1016/j.molcel.2013.07.010. [PubMed: 23932714]
22. Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet. 2006; 7:862–872. doi:10.1038/nrg1964. [PubMed: 17047685]
23. Orozco LD, et al. Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. Cell. 2012; 151:658–670. doi:10.1016/j.cell.2012.08.043. [PubMed: 23101632]
24. Song L, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome research. 2011; 21:1757–1767. doi:10.1101/gr.121541.111. [PubMed: 21750106]
25. Bennett BJ, et al. A high-resolution association mapping panel for the dissection of complex traits in mice. Genome research. 2010; 20:281–290. doi:10.1101/gr.099234.109. [PubMed: 20054062]
26. Raetz CR, et al. Kdo2-Lipid A of Escherichia coli, a defined endotoxin that activates macrophages via TLR-4. Journal of lipid research. 2006; 47:1097–1111. doi:10.1194/jlr.M600027-JLR200. [PubMed: 16479018]
27. Smale ST. Transcriptional regulation in the innate immune system. Current opinion in immunology. 2012; 24:51–57. doi:10.1016/j.coi.2011.12.008. [PubMed: 22230561]
28. Wong D, et al. Extensive characterization of NF-kappaB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. Genome biology. 2011; 12:R70. doi:10.1186/gb-2011-12-7-r70. [PubMed: 21801342]
29. Pham TH, et al. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. Nucleic acids research. 2013; 41:6391–6402. doi:10.1093/nar/gkt355. [PubMed: 23658224]
30. Jolma A, et al. DNA-binding specificities of human transcription factors. Cell. 2013; 152:327–339. doi:10.1016/j.cell.2012.12.009. [PubMed: 23332764]
31. Barski A, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007; 129:823–837. doi:10.1016/j.cell.2007.05.009. [PubMed: 17512414]
32. Wang D, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature. 2011; 474:390–394. doi:10.1038/nature10006. [PubMed: 21572438]
33. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. doi:10.1186/gb-2009-10-3-r25. [PubMed: 19261174]
34. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9:357–359. doi:10.1038/nmeth.1923. [PubMed: 22388286]
35. Hochberg YBY. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. 1995; 57:289–300.
36. Frazer KA, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature. 2007; 448:1050–1053. doi:10.1038/nature06067. [PubMed: 17660834]
37. Kirby A, et al. Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. Genetics. 2010; 185:1081–1095. doi:10.1534/genetics.110.115014. [PubMed: 20439770]
38. Kang HM, et al. Efficient control of population structure in model organism association mapping. Genetics. 2008; 178:1709–1723. doi:10.1534/genetics.107.080101. [PubMed: 18385116]
39. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. doi:10.1093/bioinformatics/bth457. [PubMed: 15297300]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

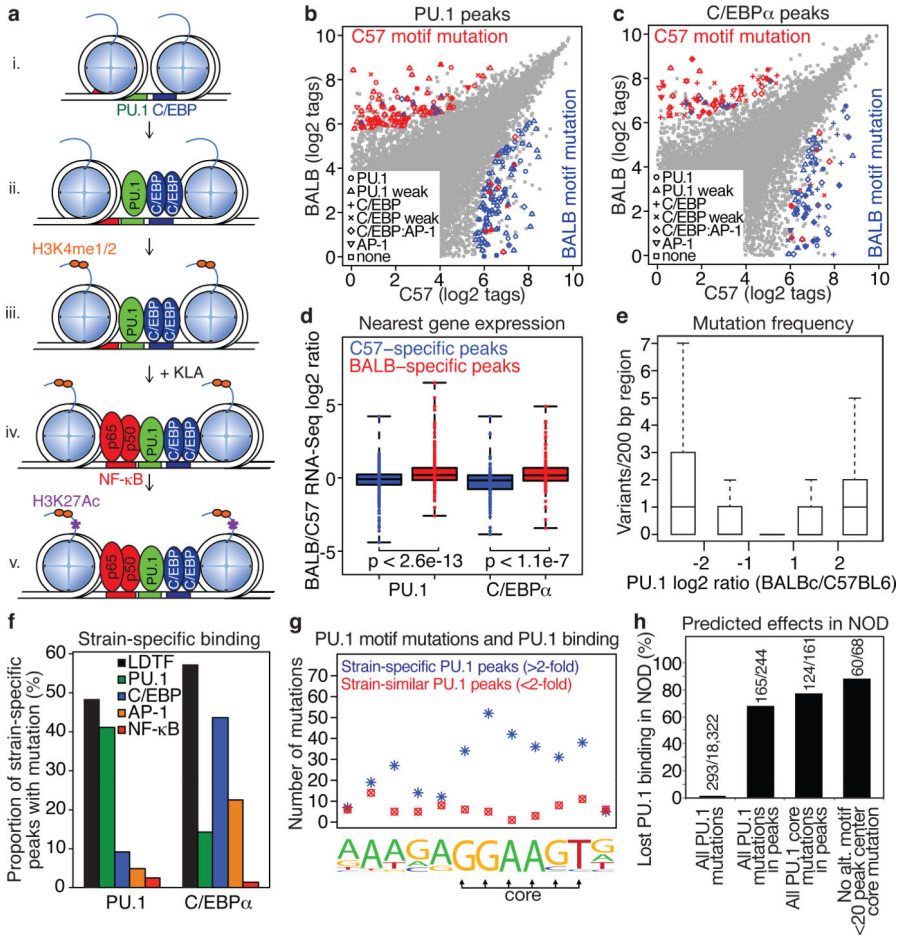


Figure 1. Genetic variation affects LDTF binding

a, Model in which LDTFs (PU.1 and C/EBPα) establish regulatory function (explained in text) **b**, **c**, ChIP-Seq-defined binding intensity for PU.1 (**b**) and C/EBPα (**c**) in resting macrophages derived from C57BL/6J (x-axes) and BALB/cJ (y-axes). Dots represent normalized tag counts in 200 bp peaks. PU.1, C/EBPα and AP-1 motifs that were mutated in one genome (distinguished by symbol; C57BL/6J = red, BALB/cJ = blue) are highlighted for peaks with strain-specific binding (4-fold, FDR ≤ $1e^{-14}$). **d,** RNA-Seq-determined expression for genes nearest to strain-specific PU.1 or C/EBPα peaks. P-values are from one-tailed t-test. **e,** Variant frequency distributions for PU.1 binding ratio bins. Box midlines (d,e) are medians, boundaries are 1st/3rd quartiles, and whiskers extend to extremes. **f**, The percentage of polymorphic, strain-specific PU.1 and C/EBPα peaks with LDTF mutations. **g,** The observed position of SNPs generating strain-specific PU.1 motifs (n = 359) underlying differential (blue) or similar (red) PU.1 binding are shown. **h.** The proportion of NOD PU.1 motif mutations that abolished PU.1 binding for each group is shown (details in **Extended Data Fig. 5**).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

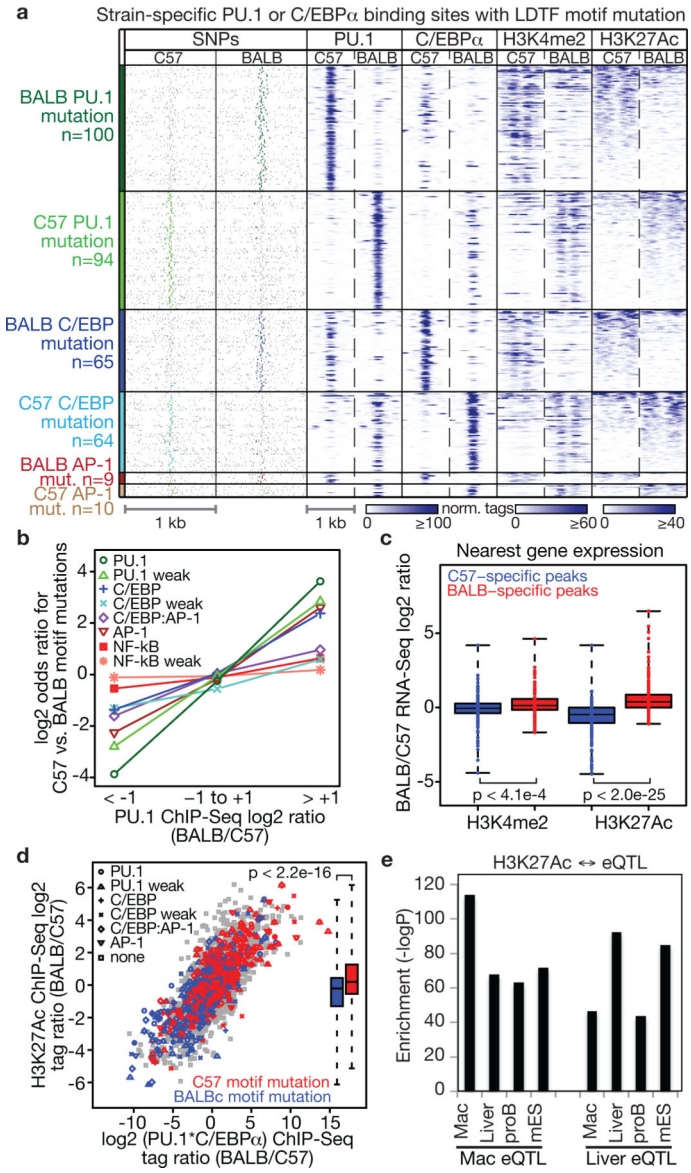


Figure 2. Genetic variation supports the LDTF collaborative binding model

a, Normalized ChIP-Seq signal at 342 loci defined by strain-specific PU.1 and/or C/EBPα binding and containing LDTF motif mutations (rows) plotted for each factor/modification (columns). Left columns display SNPs as grey dots with mutated motifs highlighted by color (LDTF mutation labels at left). **b,** Log_2 odds ratios for observing strain-specific motif mutations at strain-specific (>2-fold tag ratio, left and right bins) and similar (<2-fold tag ratio, middle bin) PU.1 peaks (details in Methods). **c,** Gene expression for genes nearest promoter-distal (>3 kb), strain-specific H3K4me2 and H3K27Ac peaks are shown (described in Fig. 1d). **d,** Normalized H3K27Ac log_2 tag ratios (1 kb, y-axis) versus log_2(PU.1xC/EBPα) tag strain ratios (200 bp, x-axis) for loci with PU.1 or C/EBPα binding. Strain-specific motif mutations are indicated by symbol and color. The distribution of H3K27Ac strain ratios stratified by strain mutations are shown (2-sided t-test). **e,**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Enrichment significance (hypergeometric distribution testing, see Methods) of H3K27Ac-modification in eQTLs from different cell types are shown.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

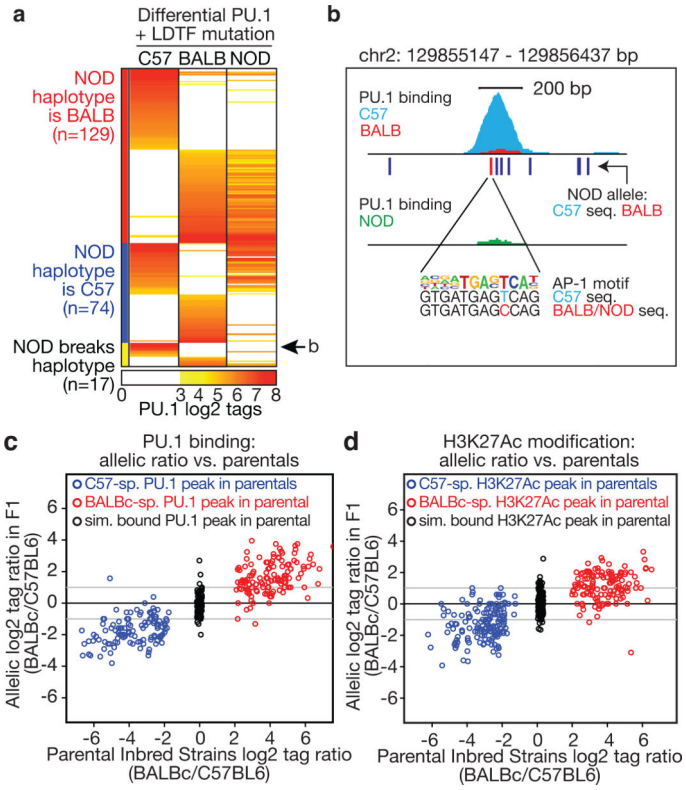


Figure 3. Validation of predicted binding and modification patterns

a, PU.1 binding at strain-specific loci are shown for C57BL/6J, BALB/cJ, and NOD/ShiLtJ mouse macrophages (columns; red = binding, white = no binding). All loci contain a strain-specific PU.1, C/EBP, or AP-1 motif. The NOD haplotype at these loci is indicated by the sidebar (red = BALB/cJ, blue = C57BL/6J, yellow = mixture). **b**, PU.1 binding, SNPs (lines), allele sharing, motif alignment and genome sequence are shown for a locus where NOD broke the C57/BALB haplotypes. **c**,**d**, Allele-specific ChIP-Seq ratios (y-axes) for PU.1 (**c**) and H3K27Ac (**d**) in CB6F1/J hybrid macrophages versus ChIP-Seq reads in parental strains (BALBc/C57 $\log_2$ ratios; x-axes) are shown for strain-specific peaks (blue = C57BL/6J, red = BALB/cJ-specific) and similarly-bound peaks (black) as defined by parental data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

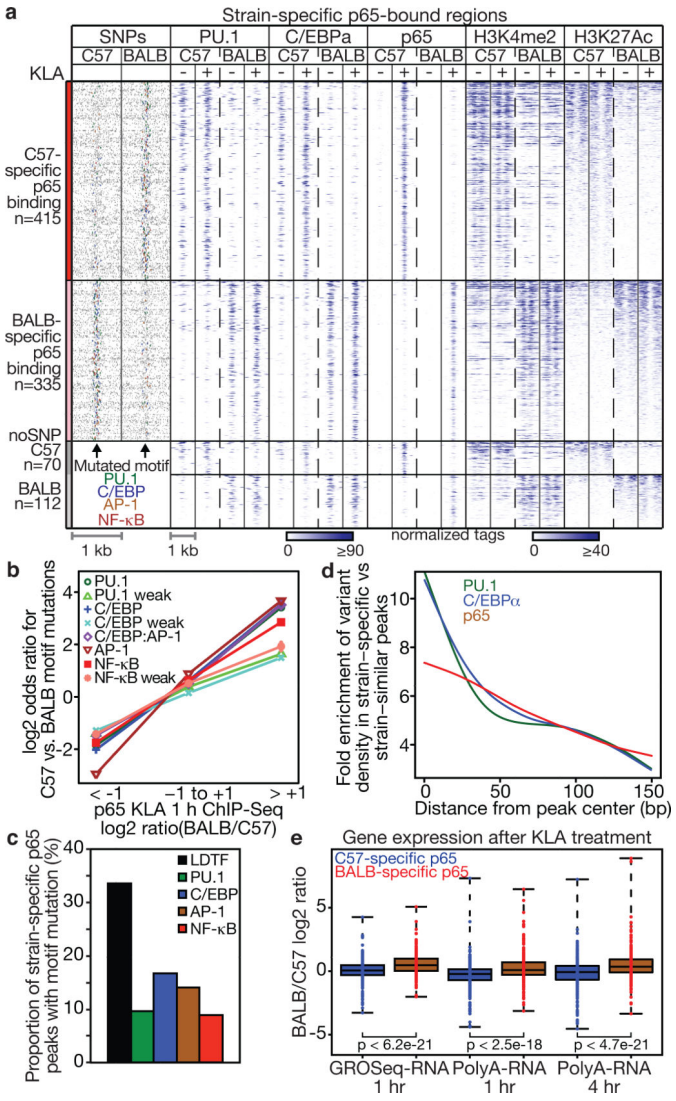


Figure 4. RelA/p65 binding is largely determined by LDTF binding

a, Strain-specific p65-bound regions were segregated into rows according to the bound strain (colored side bar). Binding/modification is shown with and without 100 ng/ml KLA treatment (−/+, 3rd header row). As in Fig. 2a, SNPs are indicated by grey dots and mutated motifs are highlighted by color (labeled beneath). **b**, The $\log_2$ odds ratio for observing strain-specific mutations are shown for bins of p65 binding as described in Fig. 2b. **c**, The percentage of polymorphic, differentially bound p65 loci harboring LDTF or NF-κB motif mutations is shown. **d**, The ratio of variant counts in strain-specific versus strain-similar peaks (y-axis) are shown relative to the peak centers for PU.1-, C/EBPα-, and p65-bound peaks in 10 bp bins (x-axis), smoothed using cubic spline. **e**, The relative amount of transcription (GRO-Seq) and mRNA production between strains after KLA treatment at the nearest gene to strain-specific p65 loci is show**n**. P-values are from one-tailed t-test.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

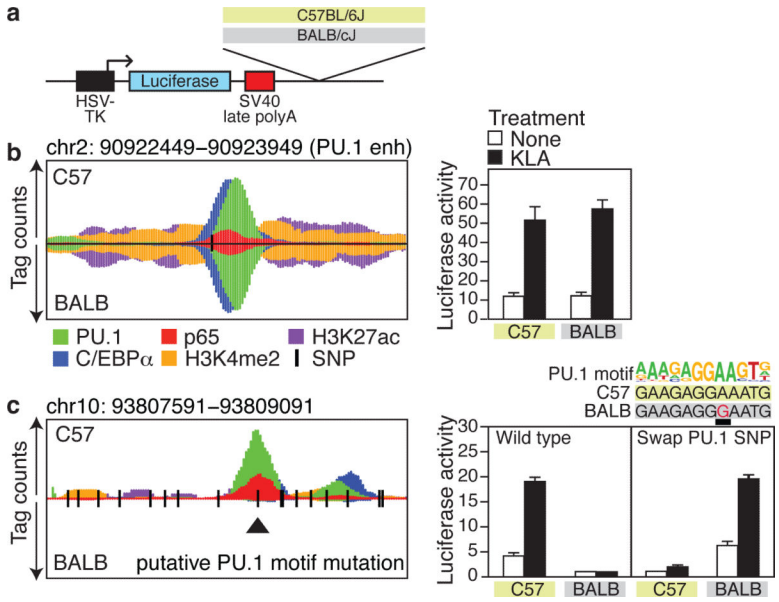


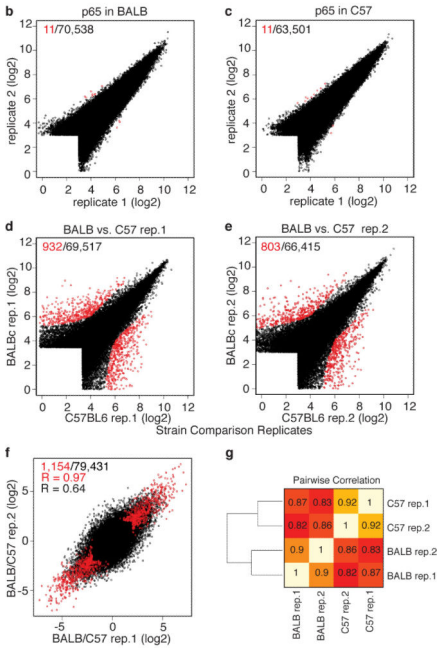
Figure 5. Validation of strain-specific enhancer activity and causal variants

a, Enhancer reporter schematic. One kb enhancer-like fragments were cloned downstream of an HSV-TK-luciferase reporter gene and tested for basal and KLA-inducible transcriptional activity in RAW264.7 macrophages. **b,** Genomic features (left) and regulatory activities (right) of the strain-similar PU.1 -14 kb enhancer positive control from C57BL/6J and BALB/cJ-derived macrophages. (**Extended Data Fig. 9-10** show all 33 loci tested). In the left panel the horizontal midline represents the 1 kb stretch of cloned DNA and SNPs are indicated with vertical black lines. ChIP-Seq tag pile-ups are shown for PU.1 (green), C/EBPα (blue), p65 (red), H3K27Ac (purple), and H3K4me2 (orange) for C57BL/6J (above midline) and BALB/cJ (below midline) with identical scales after KLA treatment (100 ng/ml, 1 h). **c,** Representative example of a strain-specific locus and the effect of a single base pair swap at the indicated PU.1 motif SNP on enhancer activity. See **Extended Data Fig. 10b,c** for additional examples and allele-swapping controls. Bargraphs plot mean ± s.d.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a

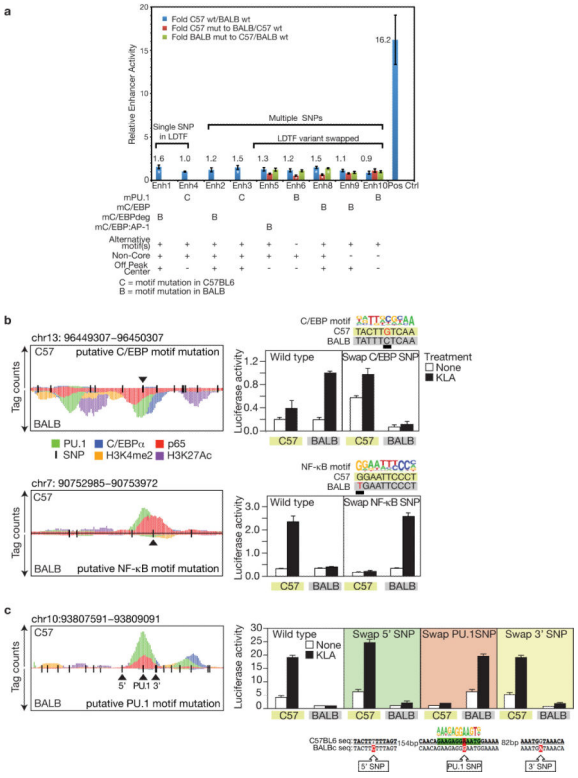
Peaks identified	Strain-similar	C57BL/6J-specific	BALB/cJ-specific	Total
H3K4me2	46,223 (99.4%)	139 (0.3%)	124 (0.3%)	46,486
H3K27Ac	22,050 (98.4%)	201 (0.9%)	168 (0.7%)	22,419
PU.1	81,658 (99.4%)	232 (0.3%)	264 (0.3%)	82,154
C/EBPα	54,611 (99.5%)	143 (0.3%)	120 (0.2%)	54,874
p65 KLA 1h	68,585 (98.7%)	485 (0.7%)	447 (0.6%)	69,517



Extended Data Figure 1. ChIP-Seq data characteristics

a, Summary of ChIP-Seq features identified. The number of ChIP-seq regions/peaks identified in untreated primary thioglycolate-elicited macrophages are tabulated for H3K4me2, H3K27Ac, PU.1 and C/EBPα. Peaks for p65 were quantified in macrophages treated with 100 ng/ml KLA for 1 hr. Unless otherwise noted, modification and binding were considered strain-specific at ≥4-fold difference between strains in sequenced tags and the FDR was <1e-14 based on Poisson cumulative distribution testing and Benjamini & Hochberg correction. **b-e,** Reproducibility and strain-specific binding. Two separate pools of thioglycolate-elicited macrophages from mice from each strain (C57BL/6J, BALB/cJ) were treated with KLA for 1 hour. ChIP-seq for p65 was performed separately on each pool (see Methods). The number of normalized sequencing tags at the union of peaks identified in the indicated experiments is shown. Peaks highlighted in red were deemed experiment-specific using criteria applied throughout this study (4-fold, and FDR<1e-14 from the cumulative Poisson distribution and Benjamini and Hochberg FDR estimation). The number of experiment-specific peaks is indicated (red) relative to the total number of peaks (black). **f,** Comparison of the p65 $\log_2$ peak tag ratio between strains and experimental sets for all peaks (black), highlighting experiment-specific peaks (red) identified in either (d) or (e). **g,** Heat map showing pairwise correlation between all p65 experiments. Pearson correlation coefficients are given for each comparison.

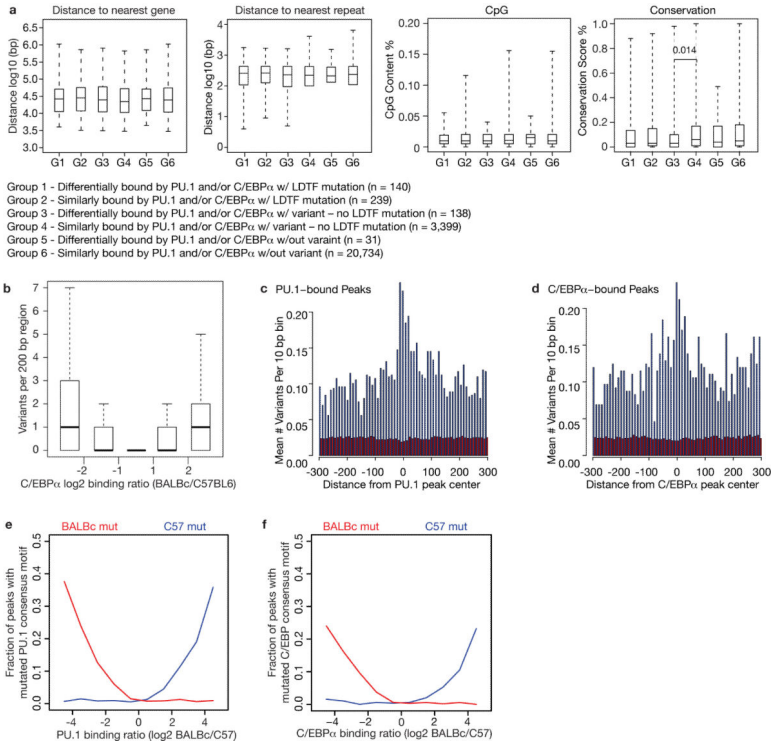
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 2. Strain-specific LDTF binding correlates with variant density and location in LDTF motifs but not with genomic context

a, Genomic features do not distinguish between strain-similar and strain-specific LDTF binding. Peaks were restricted to promoter-distal peaks (> 3 kb to gene start sites). Genomic features (distance to nearest gene, distance to nearest repeat, CpG content, conservation score) were compared among three pairs of strain-similarly bound and strain-specifically bound PU.1 and/or C/EBPα loci (listed as Group 1 – Group 6). Box midlines are medians and boundaries are 1st and 3rd quartiles. Whiskers extend to the extreme data points. CpG content and conservation were quantified in 1 kb regions centered on the LDTF peak. P-values from 2-sided t-test are given if below 0.05. **b,** Strain-specific C/EBPα binding occurs in regions with increased variant density. ChIP-Seq tag counts in 200 bp peak regions were stratified into 5 bins according to $\log_2$ ratios of peak tag counts in BALB/cJ versus C57BL/6J (x-axis, $\log_2$ ratio) and the variant density distributions are shown per bin. **c,d,** Variant density distribution in strain-specific peaks. Mean variant densities within 10 bp bins relative to ChIP-Seq peak centers in strain-similar (red) or strain-specific peaks (blue). **e,** Strain-specific PU.1 binding correlates with mutations in their respective motifs. PU.1 motif mutations were quantified in PU.1-bound regions and plotted against the logarithmic ratio of PU.1 peak tag counts in each strain (binding ratio) (x-axis). The frequency of motifs that were mutated in BALB/cJ are plotted in red and those mutated in C57BL/6J in blue. **f,** The analogous relationship as shown in **e** for PU.1 is plotted for C/EBP motif mutations versus C/EBPα strain binding ratio.

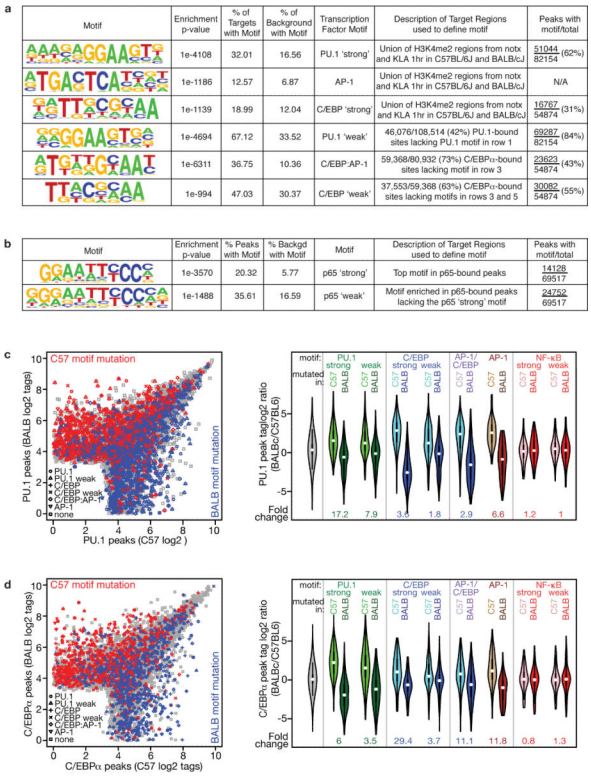
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 3. Strain-specific PU.1 and C/EBPα binding correlates with strain-specific LDTF motifs

a, Top and degenerate motifs enriched in H3K4me2 and PU.1 or C/EBPα ChIP-Seq peaks. **b,** NF-κB consensus and degenerate motifs enriched in p65 ChIP-Seq peaks. These motifs were used to query individual genome sequences and identify strain-specific motifs in subsequent analysis. Degenerate/'weak' motif occurrence numbers for a given factor include ChIP-Seq peaks containing 'strong' motifs. Position weight matrices and log-odds score thresholds for each motif are given in Supplementary Table 1. **c,** Mutations in LDTF motifs affect PU.1 and **d,** C/EBPα binding. The left panels show scatterplots for the ChIP-Seq-defined binding of PU.1 (c) and C/EBPα (d) between C57BL/6J (x-axes) and BALB/c (y-axes). Strain-specific motifs were queried within 100 bp of each peak position. Red symbols designate binding events at loci where a polymorphism mutated a C/EBP, PU.1, or AP-1 motif in the C57BL/6J genome, whereas the motif was intact in the BALB/cJ genome. Blue points highlight mutations in these motifs in the BALB/cJ genome only. Violin plots in the right panels show the effect of each motif mutation (along x-axes: PU.1, C/EBP, AP-1 and NF-κB) on the ratio of PU.1 (c) and C/EBPα (d) binding between mouse strains, (y-axes: positive values = BALB/cJ-specific, negative values = C57BL/6J-specific). Tag ratio distributions for peaks overlapping C57BL/6J motif mutations are on the left (light colors), those for peaks overlapping BALB/cJ motif mutations are on the right (dark colors). The fold-difference between mean binding ratios is indicated under the pair of distributions for each motif. The grey distribution indicates PU.1- or C/EBPα- bound loci not overlapping strain-specific motifs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

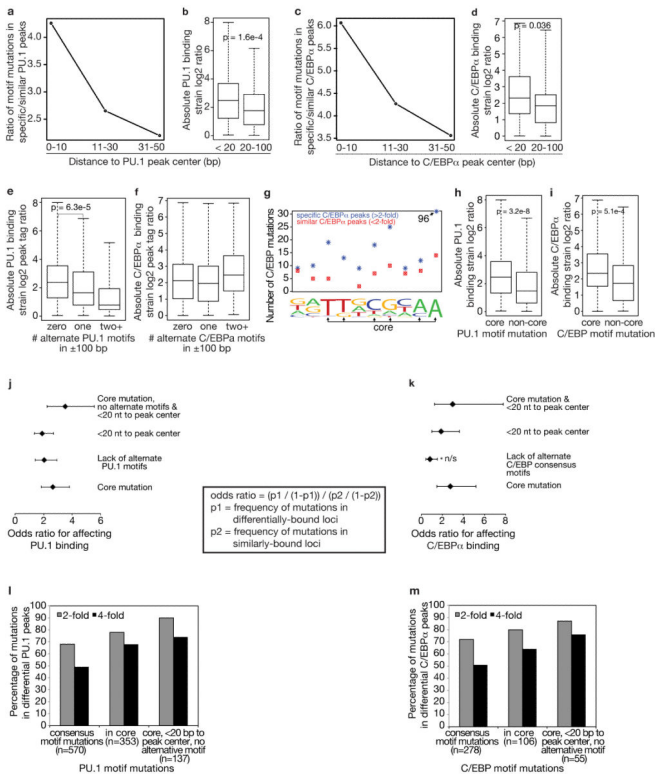


Extended Data Figure 4. Effects of cognate motif distance from peak center, variant position within a motif and the presence of alternative motifs on strain-differential binding of PU.1 and C/EBPα

a-d, PU.1 and C/EBP motif mutations near the experimentally derived peak center are associated with impaired binding. **a,c,** The ratios of the frequencies of variant-containing motifs at the given distances from strain-differentially vs. strain-similarly bound peak centers (>2-fold vs. <2-fold tag count ratio) for 570 PU.1 (**a**) and 278 C/EBP (**b**) variant-containing motifs are shown, respectively. **b,d,** The distribution of absolute strain peak tag count ratios of peaks whose center is at the given distances from mutated PU.1 (**b**) or C/EBP (**d**) motifs. Box midlines are medians and boundaries are 1st and 3rd quartiles. Whiskers extend to the extreme data points. P-values are from 2-sided t-test. **e,f,** Effects of alternative PU.1 and C/EBP motifs and core mutations on binding. The number of non-mutated, "alternate", PU.1 or C/EBP motifs in the strain with a PU.1 or C/EBP motif mutation were counted and the absolute respective PU.1 or C/EBPα $\log_2$ strain binding ratio is shown. **g,** Defining the C/EBP motif core by comparing differential vs. similar C/EBPα binding. Sequence variants within C/EBP motifs located in loci devoid of alternate C/EBP motifs (n = 178) were counted according to whether they were in differential (blue) or similar (red) C/EBPα-bound peaks. **h,** The distribution of PU.1 binding strain $\log_2$ ratios (x-axis) is shown for PU.1 mutations located in the PU.1 core and non-core nucleotides (defined in Fig. 1g). **i,** The C/EBPα binding strain $\log_2$ ratio is shown for C/EBP core and non-core mutations as defined in g. **j,k,** Motif mutations predominately occur at differentially bound loci. The odds ratios (x-axis; equation shown in box) describing the relative effect of the indicated characteristics of mutated motifs on differential binding relative to similar binding are shown for PU.1 (**j**) and C/EBPα (**k**). Whiskers show 95% confidence intervals. **l,m,** The

percentage of respective motif mutations consistent with altered PU.1 (**l**) and C/EBPα (**m**) binding are shown for the indicated categories of motif mutations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

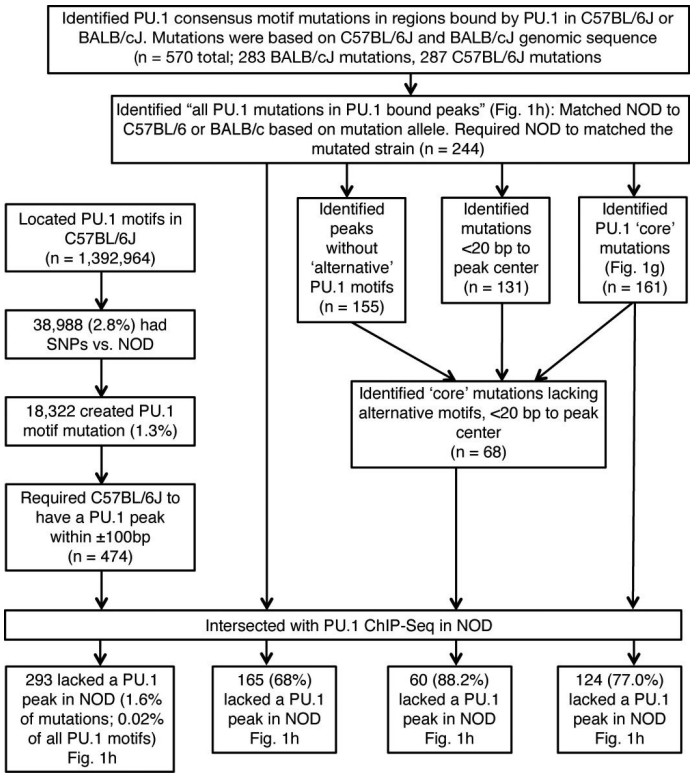


Extended Data Figure 5. Analysis pipeline for predicting functional PU.1 mutations in NOD

Data is shown in Fig. 1h.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

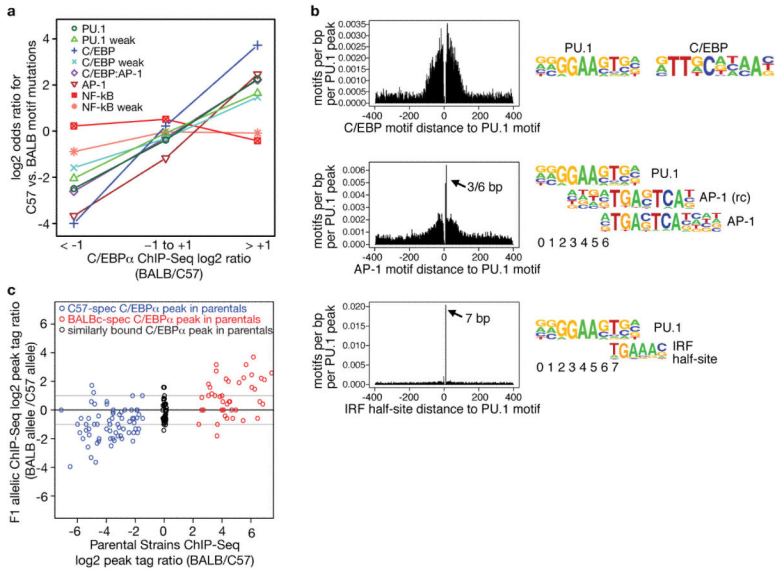
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 6. LDTF motif mutations are enriched at strain-specific C/EBPα-bound loci relative to strain-similar loci

a, The $\log_2$ odds ratio for observing a C57BL/6J- versus BALB/cJ-specific mutation in the indicated for three bins of C/EBPα binding ratios: similar (middle bin), or strain-specifically C/EBPα bound (left and right bins). Details in the Methods section. **b,** Collaborative binding is largely not mediated by direct protein-protein interactions. 14,199 loci bound by PU.1 and C/EBPα were centered on the PU.1 weak motif (0 on x-axes) and cumulative instances of C/EBP and AP-1 motifs were plotted at each position relative to the central PU.1 motif. Interferon response factor (IRF) half-sites are plotted as control for a factor that requires direct protein-protein interactions with PU.1 for DNA binding. The motifs in each comparison showing overlapping sequence and base pair distances are indicated to the right. Peak distances from the central PU.1 motif are indicated in the histograms. 'rc' in b stands for reverse complement. **c,** Allele-specific C/EBPα binding in F1 heterozygotes is similar to binding in homozygous parental strains. C/EBPα ChIP-seq reads from CB6F1/J Hybrid F1 macrophages were mapped with no mismatches to both parental genome sequences to identify allele-specific reads. C/EBPα $\log_2$ peak tag ratios between the parental strains (BALB/cJ vs. C57BL/6J) are shown on the x-axis and the $\log_2$ ratio of allele-specific reads in the F1 Hybrids are shown on the y-axis (BALB/cJ allele vs. C57BL/6J allele). C57BL/6J-specific C/EBPα regions are blue, BALB/cJ-specific C/EBPα regions are red, and strain-similar C/EBPα regions are black. Strain-specific or similar regions were defined from parental data.

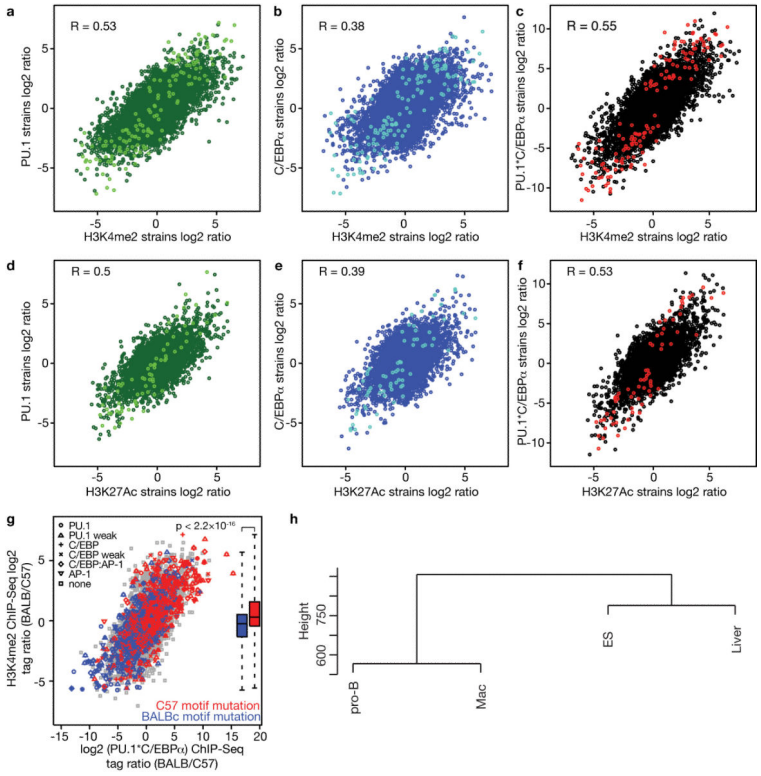
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 7. Strain-specific epigenetic marks correlate with LDTF binding, and LDTF mutations segregate with altered H3K4me2 deposition

a-f, Strain-specificity of LDTF binding and epigenetic marks. The relative amount of H3K4me2 (**a-c**) and H3K27Ac (**d-f**) between C57BL/6J and BALB/cJ (x-axes) is highly correlated with the amount of PU.1, C/EBPα, or the product (PU.1*C/EBPα) bound. The $\log_2$ ratios of the peak tag counts for PU.1, C/EBPα and PU.1*C/EBPα in each strain is shown relative to the $\log_2$ of the peak tag count ratios for H3K4me2 or H3K27Ac, respectively. Loci containing strain-specific LDTF motifs in a differentially PU.1 or C/EBPα bound peak are highlighted. Correlation coefficients (Pearson) are indicated for each comparison. **g,** LDTF mutations segregate with altered H3K4me2 deposition. The $\log_2$ of the ratio of the product of the normalized peak tag counts for PU.1 and C/EBPα in 200 bp in each strain (x-axis) is compared to the $\log_2$ H3K4me2 peak tag ratio in 1 kb (y-axis) for loci containing at least a PU.1 or C/EBPα peak. Strain-specific LDTF motif mutations are indicated by the designated symbols and colored by the mutated strain (C57BL/6J red, BALB/cJ blue). The distribution of H3K4me2 strain ratios stratified by corresponding LDTF strain mutations are shown to the right with p-value from a 2-sided t-test. **h,** Relationships between H3K27Ac patterns in different cell types. Hierarchical clustering of H3K27Ac-positive regions as determined by ChIP-Seq and analysis with HOMER. The number of ChIP-seq tags in each of the 86,264 H3K27Ac-marked regions used for comparison with eQTL data in Fig. 2e that were detected in at least one cell type were clustered using Euclidean distance.

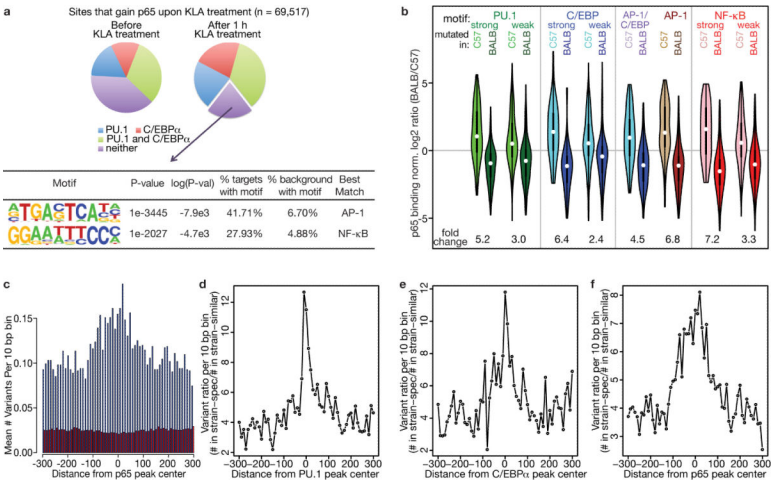
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 8. LDTFs prime the p65 cistrome

a, The 69,517 regions that gained p65 in C57BL/6J upon KLA treatment were analyzed for binding of PU.1 and C/EBPα without and with KLA treatment as shown in the pie charts. Loci not bound by PU.1 or C/EBPα after KLA treatment were analyzed by *de novo* motif finding. The most enriched motif was AP-1 and the second-most enriched motif was NF-κB. **b,** Violin plots of the p65 strain ratios of mean-normalized p65 binding for p65-bound peaks stratified by motifs mutations present in either BALB/cJ or C57BL/6J. Mutated motifs included PU.1 (strong and weak), C/EBP (strong and weak), C/EBP:AP-1, AP-1, and NF-κB. The effect on p65 binding per group is shown by comparing the mean-normalized p65 tag binding ratio along the y-axis ($\log_2$(BALB/cJ / C57BL/6J), positive values = BALB/cJ-specific, negative values = C57BL/6J-specific). White circles indicate the distribution means, and the average fold change associated with C57BL/6J-mutating and BALB/cJ-mutating SNPs in the respective motifs is given beneath. One-sided t-test p-values between each pair of distributions ranged from 1e-29 to 1e-14. **c,** Variant density in strain-specific and strain-similar p65 peaks. Mean variant density within 10 bp bins relative to p65 ChIP-Seq peak centers in strain-similar (red) or strain-specific peaks (blue). **d-e,** The variant density distribution in strain-specific p65 peaks is broader than those for PU.1 or C/EBPα. Fold enrichment of variant densities in strain-specific relative to strain-similar peaks (y-axes) for PU.1 (**d**), C/EBPα (**e**) and p65 (**f**) are shown relative to the peak centers (x-axes). Ratios plotted in d and e are from data in Extended Data Fig. 2c,d, respectively.

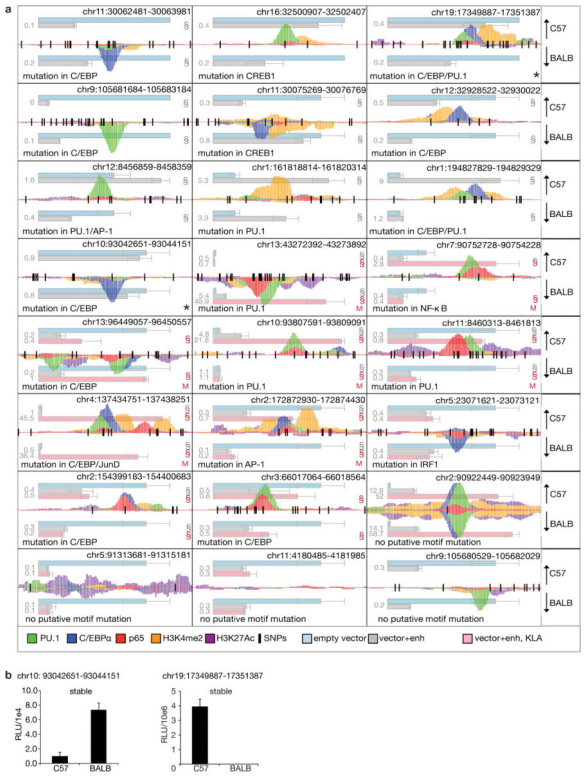
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 9. Validation of strain-specific enhancer activity

a, Enhancer activity in transient reporter assays correlates with strain-specific LDTF and p65 binding. Luciferase assay results for 24 loci (20 strain-specific enhancers with strain-specific motifs, 1 positive control with strain-similar enhancer activity (row 7, column 3), 2 negative controls lacking enhancer activity in both strains (row 8, columns 1 and 2), and 1 strain-specific enhancer lacking a strain-specific motif (row 8, column 3) in transiently transfected RAW 264.7 cells 48 hours after transfection. Each 1 kb locus is represented by the horizontal midline within a box (see Fig. 5). ChIP-seq tag pileups are shown for PU.1 (green), C/EBPα (blue), p65 (red), H3K27Ac (purple), and H3K4me2 (orange) for C57BL/6J (above midline) and BALB/cJ (below midline) with identical scales. Binding/modification data are shown after treatment with 100 ng/ml KLA. Vertical black lines indicate SNP locations. Horizontal bars indicate average luciferase (enhancer) activity of the empty vector (blue, no enhancer), activity of a locus cloned from either strain in grey C57BL/6J (above) and BALB/cJ (below) under basal conditions, or after overnight stimulation with 100 ng/ml KLA (pink). Luciferase values from transiently transfected cells were normalized to the activity measured for a co-transfected UB6 promoter-β-Gal reporter construct. Empty vector values were scaled to 0.5 for the first four loci and to 1 for the remaining loci. Error bars show standard deviations calculated from 3 biological replicates, average values are indicated next to each bar. Experiments were replicated at least 3 times. Significant strain-specific enhancer activity is indicated by § (grey without treatment, red after KLA treatment, one-tailed t-test, p-value < 0.05). **b,** Chromatinization is necessary for the strain specificity of a subset of enhancers. RAW264.7 cells were stably transfected with the two constructs harboring the loci that showed strain-specific binding but lacked strain-specific enhancer activity in transient reporter assays (row 4, column 1 and row 1, column 3, marked by *). Luciferase activity measured in lysates of stably transfected cells was normalized to total protein content.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 10. Motif analysis identifies causal SNPs in enhancers

Regions of ~1 kb size centered on PU.1 or C/EBPα ChIP-Seq peaks of similar tag count in C57BL/6J and BALB/cJ (< 2-fold difference) that contain a variant in a motif for the respective factor within 100 bp of the peak center were cloned into a luciferase reporter plasmid containing a minimal HSV-TK derived promoter. Three independent transient transfection experiments were performed in RAW264.7 cells, with triplicate transfections of each construct. Where indicated, variant nucleotides in a motif were mutated to that present in the other strain, and the resulting enhancer activity was scored relative to the wild-type allele. Shown are the ratios of the normalized luciferase activity of the C57BL/6J vs BALB/cJ alleles from a representative experiment. Luciferase values from transiently transfected cells were normalized to the activity measured for a cotransfected UB6 promoter β-Gal reporter construct, error bars represent derived standard deviations calculated by Gaussian error propagation. Constructs exhibiting significantly different activity ratios in two out of three experiments as determined by two-sided t-test ($p < 0.05$) are marked with a star. Strain and motif mutated by a variant are indicated below. In the table below, plus signs indicate whether a tested enhancer contains an alternative motif for the same factor, a variant at a motif position that is not located at a motif core as defined in Fig. 1g and Extended Data Fig. 4g, or a variant in a motif located less than 20 bp away from the peak center. Characteristics of the loci and primer sequences are in Supplementary Table 3. **b,** Identifying causal variants by motif analysis. Left panels show the ChIP-Seq pileups and SNP locations as in Extended Data Fig 9. Right panels plot the relative enhancer reporter luciferase activities of the loci shown on the left, either in the wild type configuration or when swapping the SNP indicated by a black triangle by site-directed mutagenesis. Motifs mutated by the indicated SNPs are shown above, with the mutation underlined and in red. **c,**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To confirm that the centrally located PU.1 motif is essential for the C57BL/6J- specific activity, a 1 kb fragment of the locus from C57BL/6J or BALB/cJ was cloned into the luciferase reporter described in Fig. 5 and the effects of swapping alleles at the predicted causal PU.1 SNP and flanking control 5’ and 3’ SNPs on enhancer activity are shown. Swapping alleles at the PU.1 SNP reversed strain-specific enhancer activity whereas swapping alleles at either flanking SNP had no significant effect.